Joseph Serene
Treasurer/Publisher

Phone: (301) 209-3220
Fax: (301) 209-0844
Email: serene@aps.org

American Physical Society
One Physics Ellipse
College Park, MD 20740-3844
www.aps.org

The American Physical Society's response to the OSTP's request for information on "Public Access to Digital Data Resulting from Federally Funded Research," FR Doc. 2011-32947

Introduction

The American Physical Society (APS) was founded in 1897 with the objective to advance and diffuse the knowledge of physics. While this objective is now understood to include physics education and outreach, public affairs, scientific meetings, and international collaborations, the publication of significant advances in physics has been central to APS since 1913, when we became the publisher of the *Physical Review,* a journal founded at Cornell University in 1893. Since that time, APS physics journals have grown tremendously. We now publish ten journals: *Physical Review A-E* (each journal dedicated to a particular subject area in physics), *Physical Review Letters, Reviews of Modern Physics, Physical Review Special Topics – Accelerators and Beams, Physical Review Special Topics – Physics Education Research,* and *Physical Review X*. The last three are Open Access journals whose peer-review and other operating costs are covered by contributions or publication fees. Our other journals are available through subscriptions held by a variety of individual institutions and consortia around the world. The *Physical Review* journals and *Physical Review Letters* (our flagship journal) allow authors or their sponsoring institutions to pay an article-processing charge to have an individual article made freely available. All APS Open Access articles are available under the Creative Commons Attribution 3.0 License, and we no longer hold copyright to these articles. All APS journal content (back to 1893) was made available online by the end of 2001, making us one of the very first publishers to put our entire corpus online.

The APS journals are broadly international in scope. Only about 30% of our submissions (and published articles) come from authors within the U.S. Similarly, only about one third of our subscription revenue comes from the U.S. The remaining submissions and revenues are roughly equally divided between Europe and the Asia-Pacific region.

The APS has a long history of support for Open Access initiatives. In 1998, we became the first fully "green" publisher when we amended our copyright transfer statement to explicitly allow authors to post their manuscripts (both new and previously published) on e-print servers, such as arXiv.org, and to post PDFs of their APS-published articles on their home pages or institutions' web sites. Indeed, the recent content of one of our journals, *Physical Review D*, is essentially completely available on arXiv.org because of submissions by the authors themselves. *Physical Review Special Topics – Accelerators and Beams* was one of the earliest "gold" Open Access journals. It started publication as an Open Access journal in 1998 and is supported by contributions from accelerator laboratories around the world. Authors and readers incur no fees for this journal. In November 2009, the APS Council adopted the following statement:

> *The APS supports the principles of Open Access to the maximum extent possible that allows the Society to maintain peer-reviewed high-quality journals, secure archiving, and the Society's long-term financial stability, to the benefit of the scientific enterprise.*

In keeping with our objective, APS also recognizes the importance of making the research published in our journals as widely available as possible, even to the general public. We believe that it is essential for the general public to have access through our web site to the full, final, peer-reviewed content of all APS journals. This ensures that the public sees the official "version of record," including any updates or corrections. Thus, in July 2010, we pioneered a program that allows any U.S. public library to sign up for free subscriptions to all of our journals. After a librarian at a public library completes a simple online form, agreeing to straightforward terms and conditions, we grant access promptly (usually in one business day). Any person visiting a participating public library can access the full content of our journals dating back to 1893. We are pleased that the Library of Congress was the very first public library to sign up under this program. This program was subsequently extended to all U.S. high schools, and to date well over 500 libraries from around the country have taken advantage of this opportunity.

Finally, APS prides itself on subscription prices per article and per page that are among the lowest in the industry. We were the first publisher to introduce tiered pricing, allowing smaller institutions with little research activity to pay substantially less than the leading research institutions. Subscription prices for our highest and lowest tiers currently differ by more than a factor of two, and we continue to increase (gradually) this ratio. Our article-processing fees for Open Access articles cover the actual cost of reviewing and publishing an article (without charging for submissions not accepted for publication), plus a very small margin that supports our the education and outreach activities. Twice in the most recent decade we have actually decreased our subscription prices as our expenses decreased (most recently in 2009). When we increase our prices the primary driving forces are inflation and growth in the number of manuscripts submitted for review (by 3-5% annually for many years).

OSTP Questions

Preservation, Discoverability, and Access

(1) What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?

Not all scientific data has long-term value, but the costs of preservation (at least of "bits") are already very low for all but the very largest datasets. A number of centralized data-centers for scientific information have sprung up in recent years, some of them subject-oriented (for example PANGAEA for earth science data at http://www.pangaea.de/) and some of regional scope (for example the Australian National Data Service at http://www.ands.org.au/).

Each of these services has a commitment to preservation and public access (some may allow for an embargo period) so the first helpful federal policy in this regard would be to recognize and provide funding to support scientific data centers of these sorts in the U.S. In fact a number of them are already in place – Genbank, the Protein Databank, the NIST scientific and technical database, the National Nuclear Data Center, the OSTI DOE Data Explorer, etc. One beneficial policy would include some sort of certification process to ensure that these centers are meeting at least minimal standards for data preservation and access, coupled with continued federal support and sponsorship for the creation of centers covering new subject areas.

A second beneficial federal policy would be encouraging or requiring researchers receiving federal funding to deposit their research data in a certified data center, to ensure preservation and access. To allow for reuse, deposited data must be accompanied by sufficient context – instruments used, the meaning of data fields, geographic location, time stamps, etc. – so developing standards for that context could also be a federal role (but it is surely very field-dependent).

There may also be a need for federal funding for deposit and discovery tools, so that these processes are not a burden on researchers but do actually help to improve scientific productivity. These tools have great potential and some of that potential has been realized already in fields such as genetics; it could surely be more broadly realized to improve research productivity across many other fields.

(2) What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?

We believe that the most important step here is not to mandate any specific policy, but rather to allow the choice of a certified data center that has access policies most amenable to the stakeholders involved. If the scientists are most comfortable with a one-year embargo on public access while they work on their own data, then leaving them free to choose a data-center that allows for that embargo period should be acceptable. Private data-centers that retain copyright and provide access only to subscribers may be another option that researchers could be free to choose unless federal policy insists on public access after a period of time less than the standard copyright term.

(3) How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?

Leaving the choice of (certified) data-center to the researchers should be sufficient for this.

(4) How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?

The bias should be for preserving everything. The difficulty is going to be with extremely large (particle accelerator) or diverse (bench-top science) datasets. Some filtering may be needed but long-term stewardship costs should continue to decline over time, widening the scope of what should be preserved.

(5) How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?

All should encourage the use of (certified) data-centers and development and use of appropriate tools for data deposit and discovery. Standards for metadata (data context) should be developed by the research community involved rather than by other parties.

(6) How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?

Federal research agencies should recognize and directly fund the existing data centers, and should support the creation of new ones to cover fields poorly represented at present.

(7) What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?

Every data deposit should come with an identifying citation, possibly a DOI; citations for deposited data should be provided in the same way that citations for publications are when reports are provided on grants and other funding. A project that produced data but provides no data deposit citation should be queried for compliance on the issue.

(8) What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?

Agencies should build awareness of the research data that has been made available, including tools for discovery (providing relevant hooks for search engines like Google, for instance). Tools for data manipulation, aggregation, and conversion to different formats and other overlays may need to be developed; a commercial market for innovative applications could arise.

(9) What mechanisms could be developed to assure that those who produced the data are given appropriate attribution and credit when secondary results are reported?

Data citations should include the names of the producers, or they should be provided as part of the data context. Standards for data citation are being developed by organizations such as DataCite (http://datacite.org/).

Standards for Interoperability, Re-Use and Re-Purposing

(10) What digital data standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, Nature Genetics 29, 371) is an example of a community-driven data standards effort.

[No comment]

(11) What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?

[No comment]

(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?

[No comment]

(13) What policies, practices, and standards are needed to support linking between publications and associated data?

Standards for citation of data are the most important here - in particular the work of DataCite and assignment of DOI's (digital object identifiers) to individual citable datasets should be encouraged. Then publications can link to their associated data in much they same way they link to one another now, through their reference sections with appropriate citation practice.